

# 机器学习高级技术及应用

李乡儒

Email: xiangru.li@qq.com

School of Computer Science, South China Normal University

2021 年 9 月 10 日



华南师范大学  
SOUTH CHINA NORMAL UNIVERSITY

# 定位、目标

- 培养机器学习的高级应用技术人员，帮助学生快速建立数据驱动的研究、探索、解决问题的能力
- 模型设计、构建的能力
- 典型方法的感性认识
- 解决多变的现实问题的竞争力



# 实施方案:

- 直观、深入浅出地介绍典型机器学习模型的原理
- 简单、明了地阐明相关技术的应用要点
- 具体、实战性的应用案例、强调构建能力
- 应用问题探索: Kaggle 竞赛; 最新研究问题的探索

# Preface

- **Prerequisites** 预修课程: 程序设计语言 (建议使用 Python)

- **Reference Books** 参考书:

JV2017 Jake VanderPlas(著), 陶俊杰, 陈小莉 (译). Python 数据科学手册 (Python Data Science Handbook)。人民教育出版社, 2017. (第五章介绍传统机器学习方法与基于 scikit-learn 的应用)

LCL2019 刘长龙。从机器学习到深度学习 – 基于 scikit-learn 与 TensorFlow 的高效开发实战。电子工业出版社, 2019. (第 3 章、第 4 章、第 5 章介绍传统机器学习方法与基于 scikit-learn 的应用)

ZZH2016 周志华。机器学习。清华大学出版社, 2016.

- **内容:**

- \* 传统的代表性方法: 概括介绍基本原理、基于 scikit-learn 的应用规律、参考 [JV2017][LCL2019].

- \* 神经网络、深度学习、及其在图像识别、计算机视觉等方面的应用.



华南师范大学  
SOUTH CHINA NORMAL UNIVERSITY

# 课程特点与考核:

- 特点: 直观与原理、实践性与直接应用性

- 考核, 三项作业

- Kaggle 竞赛题目: 泰坦尼克号生还者预测      考核占比 15%, 成绩在前 30% 是及格线, 网络上有大量参考; 需提交的材料: 项目研究报告/论文、实验源码、设计与运行说明, 成绩排序证明; 提交时间: 9 月 30 日 23:59 分以前, 邮件中显示的时间为准。

<https://www.kaggle.com/c/titanic>

- Kaggle 竞赛题目: House Price Prediction      考核占比 15%, 成绩在前 25% 是及格线, 网络上有大量参考; 需提交的材料: 项目研究报告/论文、实验源码、设计与运行说明, 成绩排序证明; 提交时间: 10 月 24 日 23:59 分以前, 邮件中显示的时间为准。

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques>

- 以下研究型问题之一: 空间科学数据的回归预测、分类识别, 健康图像的分割、目标检测等; 考核占比 50% (创新性与效果 30%, 撰写 20%)。需提交的材料: 项目研究报告/论文、实验源码、设计与运行说明。

提交时间: 12 月 26 日 23:59 分以前, 邮件中显示的时间为准。

- 考勤      10%
- 课堂成绩      10% [课堂回答问题、探讨、作报告/分享]



华南师范大学  
SOUTH CHINA NORMAL UNIVERSITY

# 课程项目的提交

- 可组队协作探索, 每队上限 2 人, 以组为单位提交, 需提交课程项目报告和实现程序。由于人数是奇数, 可允许一队 3 人组成, 如多个队伍报名三人组, 随机确定通过审核者。
- 提交地址: [intelligent\\_scnu@163.com](mailto:intelligent_scnu@163.com)



# 课程项目报告的撰写

- 必须使用模版
  - ① Latex 格式的模版 (排版简单, 建议使用)
  - ② Word 格式的模版
- 用自己的语言总结, 不能记流水账, 堆砌代码和图片
- 课程项目论文模板请见 QQ 群 ([878895470](#)) 的共享文件

# 优秀资源推荐

- 手写数字 0-9 识别 MNIST 库  
<http://yann.lecun.com/exdb/mnist/>  
[http://ufldl.stanford.edu/wiki/index.php/Using\\_the\\_MNIST\\_Dataset](http://ufldl.stanford.edu/wiki/index.php/Using_the_MNIST_Dataset)  
<https://msdn.microsoft.com/zh-cn/magazine/dn745868.aspx>
- 手写字符识别 (The Chars74K dataset, Character Recognition in Natural Images)  
<http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/>  
<https://ask.julyedu.com/question/7372>
- 图像识别  
STL-10 dataset      <http://cs.stanford.edu/~acoates/stl10/>  
CAIFAR-10 dataset   <http://www.cs.toronto.edu/~kriz/cifar.html>  
ImageNet            <http://www.image-net.org/>
- 最新研究: ICPR, CVPR, ICCV, ECCV, ICIP, NIPS 上感兴趣的论文
- NIPS papers: <https://papers.nips.cc/>

